

## RAW SEQUENCE LISTING

DATE: 10/24/2001

PATENT APPLICATION: US/09/866,925

TIME: 14:35:30

Input Set : A:\3124-Z.txt

Output Set: N:\CRF3\10242001\I866925.raw

1 <110> APPLICANT: Feldmann, Richard J.  
3 <120> TITLE OF INVENTION: ALGORITHMIC DETERMINATION OF FLANKING DNA SEQUENCES THAT  
4 CONTROL THE EXPRESSION OF SETS OF GENES IN PROKARYOTIC,  
5 ARCHEA AND EUKARYOTIC GENOMES  
7 <130> FILE REFERENCE: 3124-Z  
9 <140> CURRENT APPLICATION NUMBER: US 09/866,925  
C--> 11 <141> CURRENT FILING DATE: 2001-05-30  
13 <160> NUMBER OF SEQ ID NOS: 249  
15 <170> SOFTWARE: Proprietary  
18 <210> SEQ ID NO: 1  
19 <211> LENGTH: 175  
20 <212> TYPE: DNA  
21 <213> ORGANISM: E. Coli  
23 <220> FEATURE:  
24 <222> LOCATION: (3939065)...(3939239)  
25 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =  
3197  
28 <400> SEQUENCE: 1  
29 aaaaaaatgcg cggtcagaaa attatttttaa atttcctctt gtcaggccgg aataactccc 60  
30 tataatgcgc caccactgac acggaacaac ggcaaacacg ccgccgggtc agcgggggttc 120  
31 tcctgagaac tccggcagag aaagcaaaaa taaatgcttg actctgtagc gggaa 175  
34 <210> SEQ ID NO: 2  
35 <211> LENGTH: 175  
36 <212> TYPE: DNA  
37 <213> ORGANISM: E. Coli  
39 <220> FEATURE:  
40 <222> LOCATION: (4032781)...(4032955)  
41 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =  
3308  
44 <400> SEQUENCE: 2  
45 taaatttcct ctgtgcaggc cggaataact ccctataatg cgccaccact gacacggaac 60  
46 aacggcaaac acgccgccgg gtcagcgggg ttctcctgag aactccggca gagaaagcaa 120  
47 aaataaatgc ttgactctgt agcgggaagg cgtattatgc acaccccgcg ccgct 175  
51 <210> SEQ ID NO: 3  
52 <211> LENGTH: 186  
53 <212> TYPE: DNA  
54 <213> ORGANISM: E. Coli  
56 <220> FEATURE:  
57 <222> LOCATION: (3939657)...(3941012)  
58 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =  
3204  
61 <400> SEQUENCE: 3  
62 gatgtgccca gatgggatta gctagtaggt ggggtaacgg ctcacctagg cgacgatccc 60  
63 tagctggtct gagaggatga ccagccacac tggaactgag acacgggtcca gactcctacg 120  
64 ggaggcagca gtggggaata ttgcacaatg ggcgcaagcc tgatgcagcc atgccgcgtg 180  
65 tatgaa 186  
68 <210> SEQ ID NO: 4  
69 <211> LENGTH: 186

ENTERED

70 <212> TYPE: DNA  
71 <213> ORGANISM: E. Coli

## RAW SEQUENCE LISTING

DATE: 10/24/2001

PATENT APPLICATION: US/09/866,925

TIME: 14:35:30

Input Set : A:\3124-Z.txt

Output Set: N:\CRF3\10242001\I866925.raw

73 &lt;220&gt; FEATURE:

74 &lt;222&gt; LOCATION: (3941057)...(3941609)

75 &lt;223&gt; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =

3206

78 &lt;400&gt; SEQUENCE: 4

79 gtcccccttcg tctagaggcc caggacaccg ccctttcacg gcggtaacag gggttogaat 60

80 cccctagggg acgccacttg ctggtttgtg agtgaaagtc acctgcctta atatctcaaa 120

81 actcatcttc gggatgatgtt tgagatattt gctctttaa aatctggatc aagctgaaaa 180

82 ttgaaa 186

85 &lt;210&gt; SEQ ID NO: 5

86 &lt;211&gt; LENGTH: 186

87 &lt;212&gt; TYPE: DNA

88 &lt;213&gt; ORGANISM: E. Coli

90 &lt;220&gt; FEATURE:

91 &lt;222&gt; LOCATION: (3943852)...(3944312)

92 &lt;223&gt; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =

3223

95 &lt;400&gt; SEQUENCE: 5

96 gctgaagtag gtcccaaggg tatggctgtt cgccatttaa agtggtagcg gagctgggtt 60

97 tagaacgtcg tgagacagtt cggtcctat ctgccgtggg cgctggagaa ctgagggggg 120

98 ctgctcctag tacgagagga ccggagtggg cgcactactg gtgttcgggt tgtcatgcca 180

99 atggca 186

102 &lt;210&gt; SEQ ID NO: 6

103 &lt;211&gt; LENGTH: 144

104 &lt;212&gt; TYPE: DNA

105 &lt;213&gt; ORGANISM: E. Coli

107 &lt;220&gt; FEATURE:

108 &lt;222&gt; LOCATION: (3944314)...(3944450)

109 &lt;223&gt; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =

3225

112 &lt;400&gt; SEQUENCE: 6

113 aaacagaatt tgcctggcgg ccgtagcgcg gtgggtccac ctgaccccat gccgaactca 60

114 gaagtgaaac gccgtagcgc cgatggtagt gtgggtctc cccatgcgag agtagggaac 120

115 tgccaggcat caaattaagc agta 144

118 &lt;210&gt; SEQ ID NO: 7

119 &lt;211&gt; LENGTH: 112

120 &lt;212&gt; TYPE: DNA

121 &lt;213&gt; ORGANISM: E. Coli

123 &lt;220&gt; FEATURE:

124 &lt;222&gt; LOCATION: (3944469)...(3944573)

125 &lt;223&gt; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =

3228

128 &lt;400&gt; SEQUENCE: 7

129 ggtcataaaa ccggtggttg taaaagaatt ccggtggagcg gtagttcagt cggttagaat 60

130 acctgcctgt cagcgagggg gtgcggggtt cgagtcctcg ccgttcgcgc ac 112

133 &lt;210&gt; SEQ ID NO: 8

134 &lt;211&gt; LENGTH: 57

135 &lt;212&gt; TYPE: DNA

136 &lt;213&gt; ORGANISM: E. Coli

138 &lt;220&gt; FEATURE:

139 &lt;222&gt; LOCATION: (4024935)...(4025088)

140 &lt;223&gt; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =

3301

143 <400> SEQUENCE: 8

## RAW SEQUENCE LISTING

DATE: 10/24/2001

PATENT APPLICATION: US/09/866,925

TIME: 14:35:30

Input Set : A:\3124-Z.txt

Output Set: N:\CRF3\10242001\I866925.raw

```

144 ttatcgtgcc tacaaatagt ccgaaccgta ggccggataa ggcgtttacg ccgcatac      57
147 <210> SEQ ID NO: 9
148 <211> LENGTH: 56
149 <212> TYPE: DNA
150 <213> ORGANISM: E. Coli
152 <220> FEATURE:
153 <222> LOCATION: (4032754)...(4034681)
154 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
3307
157 <400> SEQUENCE: 9
158 tgccggatgc ggcgtaaacy ccttatccgg cctacggttc ggactatttg taggca      56
161 <210> SEQ ID NO: 10
162 <211> LENGTH: 347
163 <212> TYPE: DNA
164 <213> ORGANISM: E. Coli
166 <220> FEATURE:
167 <222> LOCATION: (4038097)...(4038215)
168 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
3327
171 <400> SEQUENCE: 10
172 aaaaaatgcy cggtcagaaa attatttttaa atttcctctt gtcaggccgg aataactccc      60
173 tataatgcgc caccactgac acggaacaac ggcaaacacg ccgccgggtc agcggggttc      120
174 tcctgagaac tccggcagag aaagcaaaaa taaatgcttg actctgtagc gggaaggcgt      180
175 attatgcccg tcacaccatg ggagtgggtt gcaaaagaag taggtagctt aaccttcggg      240
176 agggcgctta ccactttgtg attcatgact ggggtgaagt cgtaacaagg taaccgtagg      300
177 ggaacctgcy gttggatcac ctctttacct taaagaagcg ttctttg      347
180 <210> SEQ ID NO: 11
181 <211> LENGTH: 347
182 <212> TYPE: DNA
183 <213> ORGANISM: E. Coli
185 <220> FEATURE:
186 <222> LOCATION: (4032754)...(4034681)
187 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
3307
190 <400> SEQUENCE: 11
191 aaaaaatgcy cggtcagaaa attatttttaa atttcctctt gtcaggccgg aataactccc      60
192 tataatgcgc caccactgac acggaacaac ggcaaacacg ccgccgggtc agcggggttc      120
193 tcctgagaac tccggcagag aaagcaaaaa taaatgcttg actctgtagc gggaaggcgt      180
194 attatgcccg tcacaccatg ggagtgggtt gcaaaagaag taggtagctt aaccttcggg      240
195 agggcgctta ccactttgtg attcatgact ggggtgaagt cgtaacaagg taaccgtagg      300
196 ggaacctgcy gttggatcac ctctttacct taaagaagcg ttctttg      347
199 <210> SEQ ID NO: 12
200 <211> LENGTH: 335
201 <212> TYPE: DNA
202 <213> ORGANISM: E. Coli
204 <220> FEATURE:
205 <222> LOCATION: (4163878)...(4165793)
206 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
3432
209 <400> SEQUENCE: 12
210 tgcgcggtca gaaaattatt ttaaatttcc tottgctcagg ccggaataac tccctataat      60
211 ggcgccaccac tgacacggaa caacggcaaa cagcgccggc ggtcagcggg gttctcctga      120
212 gaactccggc agagaaagca aaaataaatg cttgactctg tagcggaag gcgtattatg      180

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,925

DATE: 10/24/2001

TIME: 14:35:30

Input Set : A:\3124-Z.txt

Output Set: N:\CRF3\10242001\I866925.raw

```

213 cacaccacac catgggagtg ggttgcaaaa gaagtaggta gcttaacctt cgggagggcg      240
214 cttaccactt tgtgattcat gactgggggtg aagtcgtaac aaggtaaccg taggggaacc      300
215 tgcggttgga tcacctcctt accttaaaga agcgt      335
218 <210> SEQ ID NO: 13
219 <211> LENGTH: 72
220 <212> TYPE: DNA
221 <213> ORGANISM: E. Coli
223 <220> FEATURE:
224 <222> LOCATION: (2729433)...(2729505)
225 <223> OTHER INFORMATION: Chromosome = 1 Strand = negative ConnectronObjectNumber =
2218
228 <400> SEQUENCE: 13
229 cttgtcaggc cggaataact ccctataatg cgccaccact gacacggaac aacggcaaac      60
230 acgccgcgg gc      72
233 <210> SEQ ID NO: 14
234 <211> LENGTH: 43
235 <212> TYPE: DNA
236 <213> ORGANISM: H. Pylori
238 <220> FEATURE:
239 <222> LOCATION: (1062106)...(1062148)
240 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
812
243 <400> SEQUENCE: 14
244 ttttactcat aggggttttta tagttcctag cggaactaaa gca      43
247 <210> SEQ ID NO: 15
248 <211> LENGTH: 43
249 <212> TYPE: DNA
250 <213> ORGANISM: H. Pylori
252 <220> FEATURE:
253 <222> LOCATION: (1158533)...(1158575)
254 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
882
257 <400> SEQUENCE: 15
258 tagcggaact aaagcattca tcccaaacac taaagatatt tgg      43
261 <210> SEQ ID NO: 16
262 <211> LENGTH: 70
263 <212> TYPE: DNA
264 <213> ORGANISM: H. Pylori
266 <220> FEATURE:
267 <222> LOCATION: (1062106)...(1062175)
268 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
813
271 <400> SEQUENCE: 16
272 ttttactcat aggggttttta tagttcctag cggaactaaa gcattcatcc caaacactaa      60
273 agatatttg      70
276 <210> SEQ ID NO: 17
277 <211> LENGTH: 70
278 <212> TYPE: DNA
279 <213> ORGANISM: H. Pylori
281 <220> FEATURE:
282 <222> LOCATION: (1158506)...(1158575)
283 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
881

```

286 <400> SEQUENCE: 17

## RAW SEQUENCE LISTING

DATE: 10/24/2001

PATENT APPLICATION: US/09/866,925

TIME: 14:35:30

Input Set : A:\3124-Z.txt

Output Set: N:\CRF3\10242001\I866925.raw

```

287 ttttactcat aggggttttta tagttcctag cggaactaaa gcattcatcc caaacactaa      60
288 agatatttgg                                     70
291 <210> SEQ ID NO: 18
292 <211> LENGTH: 70
293 <212> TYPE: DNA
294 <213> ORGANISM: H. Pylori
296 <220> FEATURE:
297 <222> LOCATION: (1062106)...(1062175)
298 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
813
301 <400> SEQUENCE: 18
302 ttttactcat aggggttttta tagttcctag cggaactaaa gcattcatcc caaacactaa      60
303 agatatttgg                                     70
306 <210> SEQ ID NO: 19
307 <211> LENGTH: 70
308 <212> TYPE: DNA
309 <213> ORGANISM: H. Pylori
311 <220> FEATURE:
312 <222> LOCATION: (1158506)...(1158575)
313 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
881
316 <400> SEQUENCE: 19
317 ttttactcat aggggttttta tagttcctag cggaactaaa gcattcatcc caaacactaa      60
318 agatatttgg                                     70
321 <210> SEQ ID NO: 20
322 <211> LENGTH: 56
323 <212> TYPE: DNA
324 <213> ORGANISM: H. Pylori
326 <220> FEATURE:
327 <222> LOCATION: (1614783)...(1614838)
328 <223> OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectronObjectNumber =
1241
331 <400> SEQUENCE: 20
332 ttttactcat aggggttttta tagttcctag cggaactaaa gcattcatcc caaaca      56
335 <210> SEQ ID NO: 21
336 <211> LENGTH: 37
337 <212> TYPE: DNA
338 <213> ORGANISM: S. Cervesiae
340 <220> FEATURE:
341 <222> LOCATION: (802985)...(803022)
342 <223> OTHER INFORMATION: Chromosome = 4 Strand = positive ConnectronObjectNumber =
1352
345 <400> SEQUENCE: 21
346 ttatgagaag ctgtcatcga agttagagga agctgaa      37
349 <210> SEQ ID NO: 22
350 <211> LENGTH: 362
351 <212> TYPE: DNA
352 <213> ORGANISM: S. Cervesiae
354 <220> FEATURE:
355 <222> LOCATION: (876188)...(876255)
356 <223> OTHER INFORMATION: Chromosome = 4 Strand = negative ConnectronObjectNumber =
1416
359 <400> SEQUENCE: 22

```



360 attagatcta ttacattatg ggtggtatgt tggaataaaa atcaactatc atctactaac

60

10/24/01

VERIFICATION SUMMARY  
PATENT APPLICATION: US/09/866,925

DATE: 10/24/2001  
TIME: 14:35:31

Input Set : A:\3124-Z.txt  
Output Set: N:\CRF3\10242001\I866925.raw

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date